

Designing Optical Fibers: Fitting the Derivatives of a Nonlinear PDE-Eigenvalue Problem

Linda Kaufman, Seok-Min Bang*, Brian Heacock*, William Landon*,
Daniel Savacool*, Nick Woronekin*

Computer Science Department, William Paterson University, Passaic, USA
Email: lkaufmang@comcast.net

Received July 3, 2012; revised September 15, 2012; accepted October 5, 2012

ABSTRACT

When trying to fit data to functions of the eigensystem of a PDE-eigenvalue problem, such as Maxwell's equation, numerical differentiation is ineffective and analytic gradients must be supplied. In our motivating example of trying to determine the chemical composition of the layers of specialty optical fibers, the function involved fitting the higher order derivatives with respect to frequency of the positive eigenvalues. The computation of the gradient was the most time consuming part of the minimization problem. It was realized that if one interchanged the order of differentiation, and differentiated first with respect to the design parameters, fewer derivatives of the eigenvectors would be required and one could take full advantage that each grid point was affected by only a few variables. As the model was expanded to cover a fiber wrapped around a spool, the bandwidth of the linearized symmetric eigenvalue problem increased. At the heart of each of the iterative methods used to find the few positive eigenvalues was a symmetric, banded, indefinite matrix. Here we present an algorithm for this problem which reduces a symmetric banded matrix to a block diagonal matrix of 1×1 and 2×2 blocks. Fillin outside the band because of pivoting for stability is prevented by a sequence of planar transformations. Computationally the algorithm is compared to the block unsymmetric banded solver and the block positive definite symmetric band solver in LAPACK.

Keywords: Symmetric; Indefinite; Eigenvalue; Gradient; Optical Fiber Design

1. Introduction

In early 2000 the Lucent Fiber Division in Norcross, Georgia, announced the specifications for a specialty fiber the design of which, they estimated would take a year. They had software in which one inserted the widths of the various layers of a fiber and the refractive index profile, indicating the chemical composition of the layer, and the program would produce such optical properties as the dispersion and micro bendloss.

However, sitting in the audience the day of the announcement was a member of the Bell Labs group from Murray Hill, NJ, which had created a fiber design tool that treated the design process as an inverse problem- the optical properties and manufacturing constraints were input and the refractive index profile was output. The next morning he handed the Georgia group three designs that met the specification, possibly removing months from the design process. The story underlines the utility of treating designing optical fibers as an inverse problem. It also indicates one of the main difficulties in the design process; the existence of multiple

solutions. In [1] the physicist W. A. Reed characterized the search for a solution to wandering on a beach with footprints and holes.

The model used by the Murray Hill group, which included the first author, assumed a fiber that was perfectly straight, circular, and uniform along its length, so that Maxwell's equations for guided waves of the fiber could be reduced to a family in (m) of problems of the form

$$\left(\frac{1}{r} \frac{d}{dr} \left(r \frac{dx}{dr} \right) + \omega^2 \alpha^2(\omega, r) - \frac{m}{r^2} \right) x = \beta^2 x. \quad (1)$$

In (1) α is a function of ω and the index of refraction profile $n(r)$, which in some regions was piecewise linear or constant as in **Figure 1** and one wishes to determine the widths, slopes and concentrations that define $n(r)$ because they dictate the chemical composition of the various layers of the fiber. We will denote these parameters as v . Among the various users of the tool there was little agreement on the shape of the burn region near the center of fiber or whether they wanted piecewise linear or piecewise constants in the other regions. The quantity ω in (1) was a specified frequency

*The last five authors were undergraduates and were supported by NSF RUI grant 0611574.

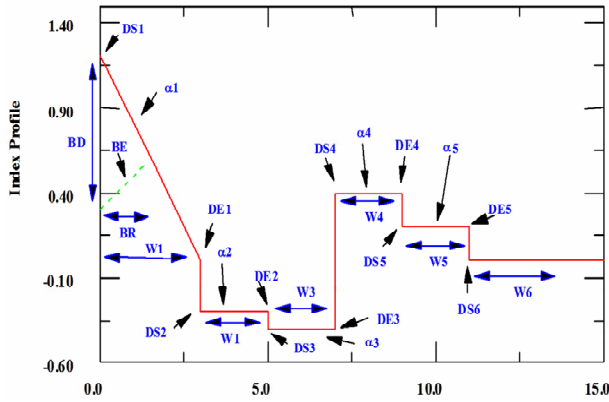


Figure 1. A typical index profile with parameters.

and m was a specified mode number. For a typical problem there might be about 20 values of ω and m might take on integer values between 0 and 2.

According to the Modified Chemical Vapor Deposition (MCVD) technique, a hollow cylinder several feet long and several inches in radius is put on a lathe and gases such as silicon dioxide, gallium and fluorine are introduced into this tube, according to the specification of the refractive index profile in $n(r)$, and these gases deposit themselves on the interior of the cylinder until it is solid. The cylinder or preform is then put in a draw tower, its end is heated, and the fiber is drawn to the specified width. The modeling process here assumes that the chemicals on a given cross section of the preform will be found in the same arrangement on the fiber.

The waveform in (1) is truncated at some radius beyond the core of the fiber and the boundary condition is expressed as the m^{th} order modified Bessel function of the second kind [2]. A finite element approach can be used to approximate (1) and converts the family of differential equations in (1) to a family of symmetric generalized tridiagonal eigenvalue problems

$$(A + s(\mu)e_q e_q^T) \mathbf{x} = \mu \mathbf{x}. \quad (2)$$

where $s(\mu)$ involves Bessel functions, A is a $q \times q$ symmetric tridiagonal matrix for a finite element discretization, e_q is the last column of the $q \times q$ Identity matrix, and M is the diagonal positive definite mass matrix which represents the inner product of the basis functions used in the finite element method. Thus the boundary condition changes the problem from finding the positive eigenvalues and corresponding eigenvectors of about 60 linear eigenvalue problems to 60 nonlinear eigenvalue problems per function evaluation.

Although solving the nonlinear eigenvalue problem is essential to designing the fiber, it is not the complete story. Depending on the ultimate use of the fiber (local area network, underwater transmissions, or to splice into an existing network to restore a degraded signal), after

the appropriate eigenvalues and eigenvectors are determined, as explained in [3], a user might wish to evaluate the dispersion given by

$$\frac{\partial^2 \beta}{\partial \lambda \partial \omega}, \quad (3)$$

the dispersion slope,

$$\frac{\partial^3 \beta}{\partial \lambda^2 \partial \omega}, \quad (4)$$

where λ is the wavelength and $2\pi c = \lambda \omega$ where the frequency ω is measured in radians per second and c is the speed of light, the effective area given by

$$2\pi \frac{\int x^2 r dr}{\int x^4 r dr}, \quad (5)$$

the Laplace Spot Size or the square of the second Peterman Radius given by $2 \frac{\int x^2 r dr}{\int \left(\frac{dx}{dr}\right)^2 r dr}$, the microbend loss,

which depends on the Laplace Spot size and the RMS Spot size = $2 \left(\frac{\int x^3 r dr}{\int x^2 r dr} \right)^{1/2}$, the macrobendloss, or the

cutoff ratio, the largest value of ω , for which there are no positive eigenvalues.

The design problem can thus be posed as a constrained nonlinear least squares problem of trying to meet target values of say the dispersion (3) and/or the effective area (5) for values of ω while keeping the width of regions nonnegative and satisfying manufacturing constraints. Because of the complexity of the eigenvalue computation and the further construction of the optical properties, optimizers tend to balk with numerical differentiation.

There are many tantalizing aspects of this problem including the determination of an appropriate grid which could change as the width parameters of the layers changed, the solution of the nonlinear eigenvalue problem, and the characterization of the nature of the multiple solutions of the inverse problem. In this paper we will concentrate on some of the ideas that were not incorporated in the original package but were developed when the first author moved to academia and was no longer in the mode of trying to meet the immediate demands of customers.

Often the solution of the eigenvalue problem required only about 20 to 30 percent of the computation time. Calculating the quantities in (3)-(5) and their derivatives with respect to the design parameters took up most of the remainder of the time. Initially one should always use symbolic and/or automatic differentiators like ADIFOR [4] to make sure that the derivative are correct, but

we found that using these techniques we were spending much time computing quantities that were zero. In Section 2 we consider ways to reduce the computation cost of the derivatives. These techniques were suitable for undergraduates like the last 5 authors and hopefully these techniques might have more general application in other inverse problems. Specifically, we were concerned about whether it was best to interchange the order of differentiation if part of the objective function was a derivative as in (3), whether it paid to use the fact that for certain parameters, like the heights of the layers, the matrix A_p was only nonzero in the portion of its rows corresponding to the grid points of the current layer, and thirdly, how could one take advantage of the fact one could write the original matrix A as a linear combination of vectors where the coefficients were dependent only on ω and the vectors were dependent only on $\mathbf{n}(r)$ —*i.e.* the variables separated.

Solving the fiber optics problem was not a one way street of technology transfer and hopefully this paper will serve to highlight some of the linear algebraic advances that were stimulated by this application. Several nonlinear eigensolvers that were applied to the problem were compared in [5], which in turn partly motivated the research in [6-10]. Each of these solvers required the solution of many symmetric indefinite tridiagonal systems. Augmenting the simple model in (1) to handle a fiber wrapped around a spool, changed the tridiagonal eigenvalue problem to a 7-diagonal eigenvalue problem with the largest elements on the main diagonal and the outermost diagonal. In [11], algorithms for solving symmetric indefinite banded tridiagonal and 5 diagonal systems are given. If the number of diagonals exceeds 5, the bandwidth could increase especially if the largest elements appeared on the outermost diagonal. The application motivated the first author of this paper, who was also a coauthor of [11] to investigate further the larger bandwidth problem. In Section 3 we discuss an algorithm for solving the symmetric indefinite banded system with bandwidths larger than 5 that was initially given in [12] and indicate how one can form a block version of the algorithm.

2. Obtaining Analytic Derivatives

In this section we look at techniques that can be used to speed up the computation of the derivatives. **Table 1** gives the percentage of the computation time used to solve the eigenvalue problem (1) and then to compute the various quantities like the dispersion slope (4) and the effective area and their derivatives with respect to the 12 design parameters. In general each run involved about solving about 1800 eigenvalue systems assuming 30 function and/or gradient evaluations per run: Each function evaluation required finding the positive eigenvalues and

their corresponding eigenvectors of about 60 systems for about 20 values of ω and 3 of m . The point of the table is that an efficient eigensolver only tackles some of the problem.

To get derivatives one can augment the system of differential equations to obtain the derivatives with respect to the design parameters but this would increase the dimension of the eigenvalue problem. A better idea would be to move immediately into the discrete domain. According to [2], if one uses a finite element method with a spacing δ , then the eigenvalue μ of (2) satisfies $\beta^2 = \mu/\delta^2 + k_{cl}^2$ where β is defined in (1) and k_{cl} is the wave number of the cladding. In general the wave number is given by $\omega n/c$.

To simplify our notation let

$$B = A + s(\mu) \mathbf{e}_q \mathbf{e}_q^T. \quad (6)$$

and let p be ω or one of the design parameters in \mathbf{v} used in defining \mathbf{n} . Modifying the classical formulation for the classical formulation for the derivative of the eigenvalues found in [13], and differentiating $B\mathbf{x} = \mu M\mathbf{x}$, we get

$$\left(\frac{\partial}{\partial p} (\mu M - A) - s_\mu \frac{\partial \mu}{\partial p} \mathbf{e}_q \mathbf{e}_q^T \right) \mathbf{x} = (\mu M - B) \frac{\partial \mathbf{x}}{\partial p} \quad (7)$$

which when multiplied by \mathbf{x}^T eliminates the right hand side of (7) giving

$$\frac{\partial \mu}{\partial p} = \mathbf{x}^T \left(\frac{\partial A}{\partial p} + \frac{\partial s(\mu)}{\partial p} \mathbf{e}_q \mathbf{e}_q^T - \mu \frac{\partial M}{\partial p} \right) \mathbf{x} / \sigma, \quad (8)$$

where $\sigma = 1 - s_\mu x_q^2$, if one stipulates that $\mathbf{x}^T M\mathbf{x} = 1$.

The formula given in (8) is basic to approximating the function value and some of the gradient of the function. The major part of the computation of the dispersion and dispersion slope required the second and third derivatives of μ with respect of ω . If we let μ' denote the derivative of μ with respect to ω , and A' the derivative of A with respect to ω , then if we differentiate (2) twice with respect to μ , multiply the result by \mathbf{x}^T and use the fact that $M' = 0$ we get the following formula to be used in the calculation of the dispersion:

$$\mu'' = (\mathbf{x}^T A'' \mathbf{x} + 2\mathbf{x}^T B' \mathbf{x}' + \alpha) / \sigma, \quad (9)$$

where $\alpha = s_{\mu\mu} \mu'^2 x_q^2$, which calls for the determination of \mathbf{x}' , which can be determined by differentiating (2) yielding

$$(\mu M - B) \mathbf{x}' = -(\mu' M - B') \mathbf{x}. \quad (10)$$

and using the fact that $\mathbf{x}^T M\mathbf{x} = 0$.

The quantity μ''' required by the dispersion slope = $\frac{\partial^3 \beta}{\partial \lambda^2 \partial \omega}$ can be derived similarly and is given by

Table 1. Percentage of time spent in various operations with 12 variables.

Operation	Function only	Derivatives also
Building A, M and their derivatives	5	16
Solving the eigenvalue problem	84	27
Derivatives of the eigensystem	0	10
Dispersion and slope	7	36
Effective area and Peterman radii	3	12

$$\mu'' = (\mathbf{x}^T A'' \mathbf{x} + 3\mathbf{x}^T B'' \mathbf{x}' + 3\mathbf{x}'^T B' \mathbf{x}'' - 3\mu' \mathbf{x}'^T M \mathbf{x}'' + \rho) / \sigma \tag{11}$$

where the quantity ρ has all the terms in $(x_q)^2 s''(\mu)$ except $\mu'' s_\mu x_q^2$.

2.1. Interchanging the Order of Differentiation

Now let us consider formulae for the derivatives of (9) and (11) with respect to a parameter p .

If one uses the convention that A_p is the derivative of the A matrix with respect to the p^{th} parameter and \mathbf{x}_p is the derivative of the eigenvector \mathbf{x} with respect to the p^{th} design parameter, then it has been shown in [14] that if θ contains all the known terms in $s(\mu)_p'' \phi^2$ then μ_p'' can be expressed several ways including

$$\mu_p'' = (\mathbf{x}^T A_p'' \mathbf{x} + 2(\mathbf{x}^T B'' \mathbf{x}_p + \mathbf{x}'^T B_p' \mathbf{x}' + \mathbf{x}^T B' \mathbf{x}_p' + \mathbf{x}_p'^T B' \mathbf{x}') + \theta) / \sigma \tag{12}$$

where θ collects the terms other than $\mu_p'' (x_q)^2 s_\mu$ in $(x_q)^2 s(\mu)_p''$ and \mathbf{x}'_p satisfies

$$(\mu M - B) \mathbf{x}'_p = -(\mu M - B)' \mathbf{x} - (\mu M - B)' \mathbf{x}_p - (\mu M - B)_p \mathbf{x}'$$

and the condition $\mathbf{x}'^T M \mathbf{x}'_p = -\mathbf{x}^T M \mathbf{x}_p'$. **Table 1** uses the formula in (12) which requires \mathbf{x}, \mathbf{x}' , and for the p^{th} design parameter \mathbf{x}_p and \mathbf{x}'_p . For 12 design parameters one must solve for 26 vectors.

Another formula for the gradient vector of the dispersion can be derived by interchanging the order of differentiation. More specifically, one begins with (7) for the p^{th} design parameter, multiplies through by \mathbf{x}^T to remove the term with the derivative of the vector and then continues to differentiate with respect to ω yielding the equation

$$\mu_p'' = (\mathbf{x}^T A_p'' \mathbf{x} + 4\mathbf{x}^T B_p' \mathbf{x}' + 2\mathbf{x}'^T B_p \mathbf{x}'' + 2\mathbf{x}^T B_p \mathbf{x}'' - \mu'' \mathbf{x}'^T M_p \mathbf{x} - 4\mu' \mathbf{x}'^T M_p \mathbf{x}' - 2\mu (\mathbf{x}'^T M_p \mathbf{x}' + \mathbf{x}^T M_p \mathbf{x}'') + \theta) / \sigma \tag{13}$$

where

$$(\mu M - B) \mathbf{x}'' = -2(\mu' M - B') \mathbf{x}' - (\mu M - B)'' \mathbf{x}, \tag{14}$$

and $\mathbf{x}'^T M \mathbf{x}'' = -\mathbf{x}^T M \mathbf{x}'$.

The second formula for μ_p'' in (13) requires $\mathbf{x}, \mathbf{x}', \mathbf{x}''$. Thus for 12 design parameters the formula in (13) requires 3 vectors. Theoretically, (12) requires at least twice as much work as (13).

The gradient of the dispersion slope (4) with respect to the p^{th} design parameter requires the derivative of μ'' with respect to that parameter. We have looked at several formulae for μ_p''' including

$$\mu_p''' = (\mathbf{x}^T A_p''' \mathbf{x} - \mu''' \mathbf{x}'^T M_p \mathbf{x} + 2\mathbf{x}^T B_p \mathbf{x}''' + \psi + 3(\mathbf{x}_p'^T B'' \mathbf{x}' + \mathbf{x}^T B_p'' \mathbf{x}' + \mathbf{x}'^T B_p' \mathbf{x}'' + \mathbf{x}_p'^T B' \mathbf{x}'') + 3\mathbf{x}'^T B_p' \mathbf{x}'' + 3\mathbf{x}^T B' \mathbf{x}_p'' - 3\mu' \mathbf{x}'^T M \mathbf{x}'' - 3\mu' (\mathbf{x}_p'^T M \mathbf{x}'' - \mathbf{x}'^T M_p \mathbf{x}'' - \mathbf{x}^T M \mathbf{x}_p'')) / \sigma \tag{15}$$

where ψ has all terms in $x_q^2 (s(\mu)_p)'''$ except $\mu_p''' (x_q)^2 s_\mu$ and

$$\mu_p''' = (\mathbf{x}^T A_p''' \mathbf{x} - \mu''' \mathbf{x}'^T M_p \mathbf{x} + 2\mathbf{x}^T B_p \mathbf{x}''' + \psi - 2\mu \mathbf{x}'^T M_p \mathbf{x}''' - 6(\mu'' \mathbf{x}'^T M_p \mathbf{x}' + \mu \mathbf{x}'^T M_p \mathbf{x}'') + \mu' (\mathbf{x}'^T M_p \mathbf{x}'' + \mathbf{x}^T M_p \mathbf{x}') - \mathbf{x}^T B_p'' \mathbf{x}' - \mathbf{x}'^T B_p' \mathbf{x}'' - \mathbf{x}'^T B_p' \mathbf{x}' - \mathbf{x}^T B_p \mathbf{x}'') / \sigma \tag{16}$$

The main advantage of (16) is that one does not have to compute $\mathbf{x}_p, \mathbf{x}'_p$, and \mathbf{x}''_p , which are the most time consuming portions of (15). It was derived by first differentiating with respect to the design parameter and then with respect to ω three times. From **Table 2** we see the advantage of using (16) and (13) over (15) and (12) just by changing the formulae. Using both sets of formulae the numerical values for the gradient of the dispersion and its slope are the same as long as the eigenvalues are accurate to full precision. Although \mathbf{x}_p is not needed in (16), it is needed to compute the gradient of the effective area if that is part of the specifications of the fiber.

The theory and methodology given in (15) for (1) can be easily extended to spooled models or full vector models and other approximations of the differential operator such as those used in [15,16], or the plane-wave

Table 2. Time (sec.) for one value of ω , one mode order, $n = 736$, 12 variables.

Eigen calculation (2 eigenvalues)	0.048
Determining x_p for all the design variables	0.017
Dispersion and slope based on (12) and (15)	0.063
Dispersion and slope based on (13) and (16)	0.033
Dispersion and slope based on (13) and (16) taking advantage of the zero structure of matrices in formulae	0.015

method [17] or a finite difference method [18].

2.2. Taking Advantage of the Structure of the Derivative Matrices

Another advantage of (16) is that all the matrices in the formulae involve derivatives with respect to the design variables \mathbf{v} and all these parameters had a limited range of support. If the p^{th} design parameter denoted the height of a profile in a layer then $M_p = 0$ and A_p, A'_p and A''_p were 0 for those grid points outside the layer. If the p^{th} design parameter denoted the width of a layer, the grid points corresponding to the layers closer to the center of the fiber were zero, but because the preform had a finite width, those grid points within the layer and further from the center contributed nonzero value to the M and A matrices.

Using the fact that some of these matrices had only a few rows that were nonzero decreased the time for constructing the matrices A_p, A'_p, A''_p , and A'''_p as well as computing the effective area. Reworking the code to take advantage of the fact that these matrices had a small range of support was not trivial. The program for building the matrices and using them processed one grid point at a time and symbolic differentiation was used to insert lines for computing the derivative with respect of \mathbf{v} . In certain parts of the code attempting to take advantage of the support grid points for each parameter meant changing data structures and inverting loops; in other portions, the code was changed so that the matrices were constructed based on layers, whose number was an input parameter, and not on grid points, whose location and number could be readjusted for every value of ω at every function value.

Lastly, as more derivatives with respect to ω were required, the formulae for the derivatives with respect to ω and \mathbf{v} of the potential $\phi(\omega, \mathbf{v}) = \omega^2 \alpha(\omega, \mathbf{n}(\mathbf{v}))$ in (1) became more complicated. But actually the potential at the i^{th} grid point $\phi_i(\omega, \mathbf{v})$ can be rewritten as

$$\phi_i(\omega, \mathbf{v}) = s_1(\omega)c_i(\mathbf{v}) + s_2(\omega)c_i(\mathbf{v})^2 \quad (17)$$

where $c_i(\mathbf{v})$ denotes the dopant concentration at the i^{th} grid point, and $s_1(\omega)$ and $s_2(\omega)$ are functions of

ω and the Sellmeier coefficients [19]. To add a new derivative of ϕ with respect to ω , entails determining new values of the derivatives of s_1 and s_2 once for all grid points and then 2 multiplications are needed for each grid point. Similarly once $\frac{\partial c_i}{\partial v_p}$ has been computed, it

can be used with the appropriate derivatives of s_1 and s_2 to form A_p, A'_p, A''_p , and A'''_p . Note for the parameters that denote the widths of the layers $\frac{\partial c_i}{\partial v_p}$ was

zero for all grid points and for the other parameters which were used to dictate the height of a grid point in a particular layer, $\frac{\partial c_i}{\partial v_p}$ was zero outside of their own

layer.

Using the separable form of (17) greatly simplifies the construction of the matrices in (12). **Table 3** gives the proportion of time spent in various sections in the final code. The big hit that was initially seen in **Table 1** for of the calculation of analytic derivatives has disappeared.

3. Fiber around a Spool-Solving Banded Symmetric Indefinite Linear Systems

For the model in (1) at each iteration one was concerned with getting the few positive eigenvalues and their eigenvectors. Usually a generalized Rayleigh quotient iteration that minimized $\|(A + s(\mu)e_q e_q^T)x - \mu Mx\|$ was effective:

Nonlinear Rayleigh Quotient

1) Determine a lower bound μ_l and an upper bound μ_u for the desired eigenvalue μ .

Set μ_0 to μ_l .

2) Set $\mathbf{x} = \mathbf{e}$, the vector of all 1's if a good approximation to the eigenvector is not available from a previous iteration.

Set j to 0.

3) Until convergence.

a) Set $B = A - s(\mu_j)\mathbf{e}_n \mathbf{e}_n^T$.

b) Solve $(B - \mu_j I)\mathbf{y} = \mathbf{x}$ and determine the inertia of $(B - \mu_j I)$.

c) If the inertia claims that μ_j is less than μ , reset μ_l to μ_j .

d) If the inertia says that μ_j is greater than μ , reset μ_u to μ_j .

e) Set $\mathbf{z} = B\mathbf{y}$.

f) Set $\alpha = \frac{\mathbf{z}^T \mathbf{y} - s'(\mu)y_n z_n}{\mathbf{y}^T \mathbf{y} - s'(\mu)y_n^2}$.

g) If $\alpha < \mu_l$, set $\mu_{j+1} = 0.9\mu_l + 0.1\mu_u$.

If $\alpha > \mu_u$, set $\mu_{j+1} = 0.9\mu_u + 0.1\mu_l$.

as follows

1) Let $\lambda^{(k)} = |a_{r,1}^{(k)}|$ = maximum element in absolute value in the first column.

2) If $|a_{1,1}^{(k)}| \geq \gamma\lambda^{(k)}$, use a 1×1 pivot to obtain $A^{(k-1)}$, decrease k by 1) and return to 1). Here γ is a scalar of about 1/3 to balance element growth.

else

3) Determine $\sigma^{(k)}$, largest element in absolute value of the r th column.

4) If $\gamma\lambda^{(k)^2} \leq |a_{1,1}^{(k)}|\sigma^{(k)}$, use a 1×1 pivot to obtain $A^{(k-1)}$, decrease k by 1) and return to 1).

else

5) Interchange the r th and second rows and columns of A^k and partition it like (21).

6) Form the Z matrix $Z = D^{-1}Y^T$.

7) For $i = 1, \dots, \min(r-3, k-r)$.

Do a sequence of planar rotations in planes $(i, r-2)$ to eliminate elements in the second row of Z and apply these transformations to the corresponding rows of Y , and $A^{(k)}$

8) Perform a 2) by 2) pivot to obtain $A^{(k-2)}$, decrease k by 2) and return to 1)

For a matrix A of n rows, the retraction algorithm outlined above using stabilized elementary transformations for the planar rotations requires between $1/2 gn^2$ and $5/4 gn^2$ multiplications compared to between gn^2 and $2gn^2$ multiplications required by unsymmetric Gaussian elimination. It requires $(2g+1)n$ elements to hold the Z matrices compared to $(3g+1)n$ operations. As reported in [12] where the planar transformations were introduced not as a preventative measure but as a way of cleaning up elements that appeared outside the original band structure, an implementation of the algorithm was faster than LAPACK's DGBTF2 [21] for nonsymmetric banded matrices. However, when g was sufficiently large so that it was advantageous to use a block structured algorithm, the retraction algorithm could not compete with LAPACK's DGBTRF.

Since the publication of [12], it was realized that it was possible to use the extra space required for the Z matrices to generate a block approach for those portions of the algorithm whenever the planar transformations were not required. By embedding the matrix in the larger $(2g+1) \times n$ array one can avoid some of the difficulties in the block symmetric codes in LAPACK which store and access only the lower triangular portion of the matrix. Similarly one can avoid a few of the problems associated with the banded codes of continually copying and retrieving a block or a triangle to a scratch space. The sketch in **Figure 2** indicates how we partitioned the matrix with the dashed lines indicating the structure of

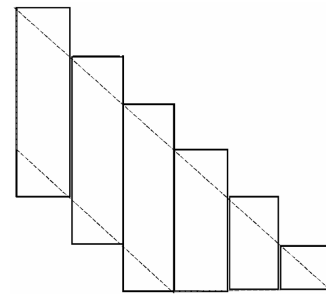


Figure 2. Computational idea of lower triangular portion of banded matrix.

the mathematical object and the sequences of rectangles the computational object. When applying transformations matrix-matrix operations could be used exclusively.

In **Figure 3** the times for the symmetric indefinite block version of the retraction algorithm is compared to DPBTRF, Lapack's positive definite banded routine block version, and DGBTRF for various bandwidths. For the retraction algorithm the block size is 16 while for that of the other algorithms, it is set at 32. ATLAS BLAS were used in FORTRAN.

In **Figure 4** the matrices all have the same bandwidth, but the outer diagonal and the main diagonal are varied to produce matrices that require various number of planar transformations. In this version, a block concept of the retraction algorithm is pursued until one hits a 2 by 2 pivot that requires planar transformations. At the cross-over point about 60 percent of the rows are involved in 2×2 blocks. It is obvious that the more transformations, the greater the time.

4. Conclusions

Although this research was directed at the fiber optic industry, we hope its impact will have a significant wider breath to cover other PDE-eigenvalue problems with design parameters. Specifically, this paper should be applicable to inverse problems where finding the gradient is absolutely necessary but seems to involve a tremendous amount of computation. We have shown that it pays to take advantage of local support of variables, to write the function in a separable form if possible and if the function is itself a derivative, to ask whether it pays to change the order of differentiation when producing the gradient.

Trying to solve the spooled fiber optics problem suggested to the first author that there are linear systems that are symmetric indefinite and banded. Her first attempts seemed to suggest that only for examples where the inertia is definitely required or the bandwidth is not large, that the retraction algorithm is useful. However, recently she has realized that the extra space required by the algorithm to store information for solving linear systems could be used to produce a viable block version.

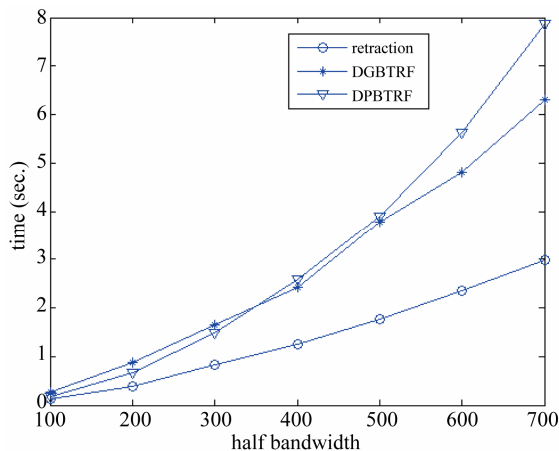


Figure 3. Time vs. half bandwidth, $n = 2000$, positive definite matrices.

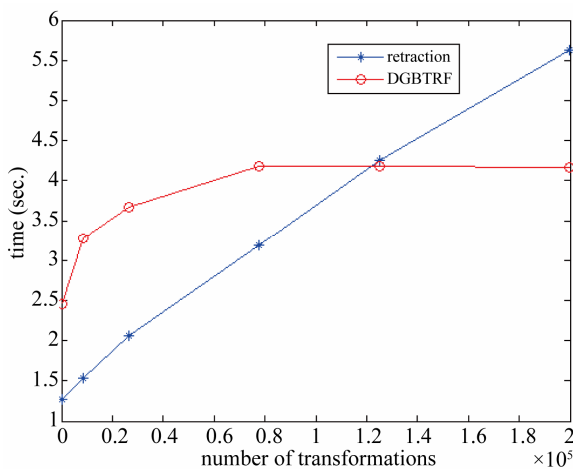


Figure 4. Time vs. number of planar transformations, $n = 2000$, $g = 400$.

5. Acknowledgements

All coauthors were supported in part by a National Science Grant for research at undergraduate institutions.

REFERENCES

- [1] W. A. Reed, "Fiber Design for the 21st Century," Optical Society of America, San Francisco, 1999.
- [2] T. A. Lenahan, "Calculation of Modes in an Optical Fiber using the Finite Element Method and EISPACK," *The Bell System Technical Journal*, Vol. 62, No. 9, 1983, pp. 2663-2694.
- [3] S. V. Kartalopoulos, "Introduction to DWDM Technology," IEEE Press, Piscataway, 2000.
- [4] C. Bischof, A. Carle, G. Corliss, A. Griewank and P. Hovland, "ADIFOR-Generating Derivative Codes for Fortran Programs," *Scientific Programming*, Vol. 1, No. 1, 1992, pp. 11-29.
- [5] L. Kaufman, "Eigenvalue Problems in Fiber Optics Design," *SIAM Journal on Matrix Analysis and Applications*, Vol. 28, No. 1, 2006, pp. 105-117. [doi:10.1137/S0895479803432708](https://doi.org/10.1137/S0895479803432708)
- [6] X. Huang, Z. Bai and Y. Su, "Nonlinear Rank One Modification of the Symmetric Eigenvalue Problem," *Journal of Computational Mathematics*, Vol. 28, No. 2, 2010, pp. 218-234.
- [7] T. Betcke, N. J. Higham, V. Mehrmann, C Schröder and F. Tisseur, "NLEVP: A Collection of Nonlinear Eigenvalue Problems," Manchester Institute of Mathematical Sciences Preprint, Manchester, 2008.
- [8] H. Voss, K. Yildiztekin and X. Huang, "Nonlinear Low Rank Modification of a Symmetric Eigenvalue Problem," *SIAM Journal on Matrix Analysis and Applications*, Vol. 32, No.2, 2011, pp. 515-535. [doi:10.1137/070779528](https://doi.org/10.1137/070779528)
- [9] H. Schwetlick and K. Schreiber, "Nonlinear Rayleigh Functionals," *Linear Algebra and its Applications*, Vol. 436, No. 10, 2011, pp. 3991-4016.
- [10] B. S. Liao, Z. Bai, L.-Q. Lee and K. Ko, "Nonlinear Rayleigh-Ritz Iterative Methods for Solving Large Scale Nonlinear Eigenvalue Problems," *Taiwanese Journal of Mathematics*, Vol. 14, No. 3, 2010, pp. 869-883.
- [11] J. R. Bunch and L. Kaufman, "Some Stable Methods for Calculating Inertia and Solving Symmetric Linear Systems," *Mathematics of Computation*, Vol. 31, No. 137, 1977, pp. 163-179. [doi:10.1090/S0025-5718-1977-0428694-0](https://doi.org/10.1090/S0025-5718-1977-0428694-0)
- [12] L. Kaufman, "The Retraction Algorithm for Factoring Banded Symmetric Matrices," *Numerical Linear Algebra with Applications*, Vol. 14, No. 3, 2007, pp. 237-254. [doi:10.1002/nla.529](https://doi.org/10.1002/nla.529)
- [13] P. Lancaster, "On Eigenvalues of Matrices Dependent on a Parameter," *Numerische Mathematik*, Vol. 6, No. 1, 1964, pp. 377-387. [doi:10.1007/BF01386087](https://doi.org/10.1007/BF01386087)
- [14] L. Kaufman, "Calculating Dispersion Derivatives in Fiber-Optic Design," *Journal of Lightwave Technology*, Vol. 25, No. 3, 2007, pp. 811-819. [doi:10.1109/JLT.2006.889648](https://doi.org/10.1109/JLT.2006.889648)
- [15] F. Brechet, J. Marcou, D. Pagnoux and P. Roy, "Complete Analysis of the Characteristics of Propagation into Photonic Crystal Fibers by the Finite Element Method," *Optical Fiber Technology*, Vol. 6, No. 2, 2000, pp. 181-191. [doi:10.1006/ofte.1999.0320](https://doi.org/10.1006/ofte.1999.0320)
- [16] K. Saitoh, M. Koshiba, T. Hasegawa and E. Sasaoka, "Chromatic Dispersion Control in Photonic Optical Fibers: Application to Ultra-Flattened Dispersion," *Optics Express*, Vol. 11, No. 8, 2003, pp. 843-852. [doi:10.1364/OE.11.000843](https://doi.org/10.1364/OE.11.000843)
- [17] S. Johnson and J. Joannopoulos, "Block-Iterative Frequency-Domain Methods for Maxwell's Equations in a Planewave Basis," *Optics Express*, Vol. 8, No. 3, 2001, pp. 173-190. [doi:10.1364/OE.8.000173](https://doi.org/10.1364/OE.8.000173)
- [18] S. Guo, F. Wu, S. Albin, H. Tai and R. Rogowski, "Loss and Dispersion Analysis of Microstructured Fibers by Finite Difference Method," *Optics Express*, Vol. 12, No. 15, 2004, pp. 3341-3352. [doi:10.1364/OPEX.12.003341](https://doi.org/10.1364/OPEX.12.003341)
- [19] J. W. Fleming, "Material Dispersion in Lightguide Glasses," *Electronics Letters*, Vol. 14, No. 11, 1978, pp.

- 326-328. [doi:10.1049/el:19780222](https://doi.org/10.1049/el:19780222)
- [20] D. Marcuse, "Field Deformation and Loss Caused by Curvature of Optical Fibers," *Journal of the Optical Society America*, Vol. 66, No. 4, 1972, pp. 311-320. [doi:10.1364/JOSA.66.000311](https://doi.org/10.1364/JOSA.66.000311)
- [21] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov and D. Sorensen, "LAPACK Users' Guide," 2nd Edition, Society for Industrial and Applied Mathematics, Philadelphia, 1995.